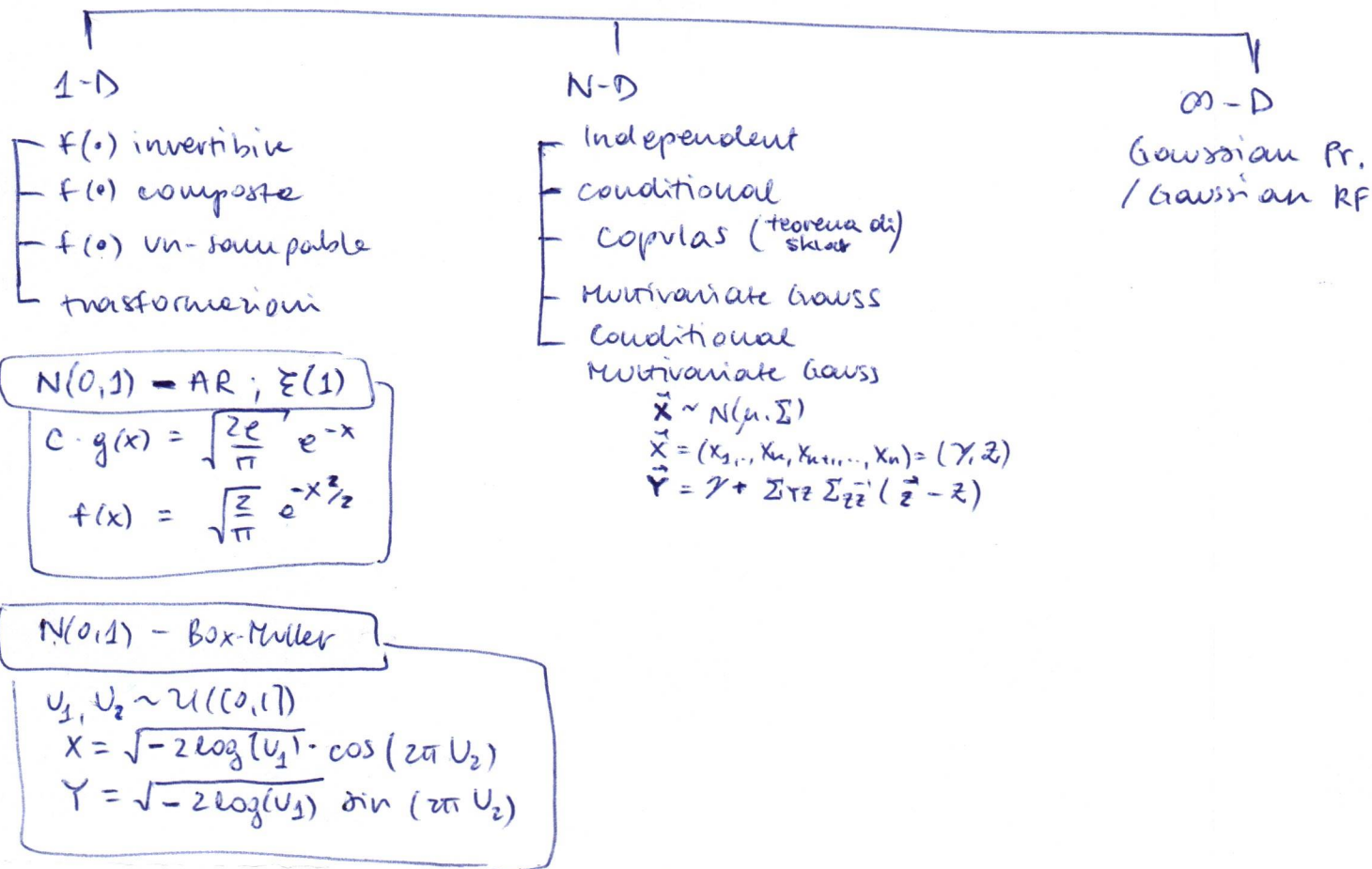


RANDOM VAR. GEN



MC : LLN : $P(\hat{\mu} \rightarrow \mu) = 1$
 TCL : $\frac{\sqrt{N}(\hat{\mu} - \mu)}{\sigma} \rightarrow N(0,1)$

$$P(\mu \in I_\alpha) \rightarrow 1 - \alpha \quad : \quad \begin{cases} \hat{I}_\alpha = \left[\hat{\mu} \pm z_{1-\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{N}} \right] \\ \hat{\sigma} = \frac{1}{N-1} \sum_{i=1}^N (z^{(i)} - \hat{\mu})^2 \\ z_\alpha : \Phi(z_\alpha) = \alpha \end{cases}$$

VARIANCE REDU.

ANTITHETIC VAR.	$\rightarrow \hat{\mu}_{AV} = \frac{1}{N/2} \sum_{i=1}^{N/2} \frac{\psi(\vec{x}^{(i)}) + \psi(Z E[\vec{x}^{(i)}] - \vec{x}^{(i)})}{2}$
IMPORTANCE SAMPLING	$\rightarrow \int \frac{(\psi(\vec{x}) f(\vec{x}))^2}{g(\vec{x})} d\vec{x} \Rightarrow g^*(\vec{x}) = \frac{ \psi(\vec{x}) \cdot f(\vec{x})}{E[\psi(\vec{x})]}$
CONTROL VARIATES	$\rightarrow \begin{cases} \text{Var}(\tilde{z}_\alpha) = \text{Var}(Z) + \alpha^2 \text{Var}(Y) + 2\alpha \text{Cov}(Z, Y) \\ \alpha^* = - \frac{\text{Cov}(Y, Z)}{\text{Var}(Y)} \end{cases}$
STRATIFICATION	
LHS	$\text{Var}(\tilde{z}_\alpha^*) = \text{Var}(Z) (1 - \text{Corr}(Z, Y)^2)$

ZAREMBA \rightarrow KOKSMA

CAPITOLO 7

- global methods: VARIANCE BASED SENSITIVITY ANALYSIS (Sobol' indices) (conditional expectations and variances (i.e. moments))
 - * monte carlo evaluation -> many evaluations
 - * analysis of variance
- local methods: SCREENING METHODS (Morris)
 1. Can we avoid moments? monte carlo is heavy computationally.
 - > Moment independent sensitivity
 2. Can we replace the original model with a surrogate /emulator?
 - > Reduced order modelling
 - > Gaussian Process Regression
 - > Polynomial chaos

In entrambi gli ultimi due casi we speed the evaluation of Sobol indices:

 - model -> Sobol indices (MC) X
 - surrogate -> Sobol indices V
- SA quantifica gli effetti of input parameters variation on outputs of interest, providing a criterium to rank the most influential input parameters
- study of how uncertainty in the output of a model ..
- l'ideale è di combinare SA e uncertainty propagation
- SA non è studiare la varianza dell'output: è saper responsabilizzare una porzione di varianza dell'output a un certo effetto/parametro/variabile aleatoria/input
- 4 tipi di input
- 5 tipi di task
- due divisioni SA: local (small input perturbation), global (analysis of the entire space)
- $Y=f(X_1, \dots, X_k)$, dove gli input sono indipendenti gli uni dagli altri, di conseguenza possono essere sampled dalla loro marginale (SE non specificato diversamente) (in realtà per introdurre i metodi supponiamo che le X_i siano indipendenti e uniformemente distr in $[0,1]$)

LOCAL

- primo approccio: based on partial derivatives (chiaramente è sbatti)
 - problematici se:
 - * il modello che genera l'output è complesso
 - * la dimensione dello spazio è grande
 - * diversi effetti interagiscono tra di loro
 - difficile esplorare tutto lo spazio, cosa che invece vorremmo
 - > se non guardiamo tutto lo spazio rischiamo di perderci pezzi (counterexample)
 - > possiamo migliorare con la sigma-normalized output derivatives
- secondo approccio: (chiaramente ha delle assunzioni di linearità e normalità, heavy)

"a proposito di scatterplot": very simple and informative way of running a SA.

 - > to summarize the results of a scatterplot we can do a simple regression
 - > se supponiamo che una regressione lineare possa essere una buona struttura allora i standardized regression coefficients sono una misura of the sensitivity of the output to the input i , in particolare i SRC² sono the fractional contribution to the variance of the model output given by the input i .

Nello stesso framework, un'altra sensitivity measure è data dal coefficiente di correlazione tra l'output e l'input i (in realtà se gli input sono indipendenti la correlazione è data sempre dai SRC). Addirittura, se il modello è lineare le SRC sono le sigma-normalized derivatives. Se il modello non è lineare allora i due così non coincidono e SRC è più robusto.

RECAP: SRC are based on a linear regression model for the output function, and are computed from the least square regression analysis applied to the output of a MC simulation. Their validity as a measure of sensitivity analysis is conditional to the degree to which the regression model fits the data, i.e. R^2 (coefficiente di determinazione, usato to assess the goodness-of-fit of a linear model)
- terzo approccio: SCREENING METHODS (non è basato sulle partial derivatives, o almeno non in un modo analitico)
 - Il problema delle derivate parziali è che ci mettiamo tanto a calcolarle e farlo in ogni punto diventa sbatti. However, if at each point we analyze the incremental ratio instead of the derivatives, we end up with a good method: SCREENING METHOD
 - L'idea è di perform parameter variation one-at-the-time, where each input is varied while fixing the others. Ci sono diversi modi di procedere: una è la ELEMENTARY EFFECTS (EE) o Morris method, dove il numero di esperimenti è dello stesso ordine del numero di input.
 - EE identifica gli inputs non influenti nei casi in cui il modello matematico è computazionalmente costoso oppure nel caso in cui si hanno tanti inputs.
 - EE classifies gli inputs in tre gruppi: negligible effects, large linear effects without interactions and large non-linear and/or interaction effects.
 - Il metodo (EE) consiste nel discretizzare l'input space per ciascuna variabile e poi performare un certo numero di OAT (one-at-time) design experiments. I design e le direzioni di variazione sono scelti casualmente.
 - METODO:
 - * discretizziamo the parameters space in ciascuna X_i (creiamo una griglia dei possibili valori che può assumere il vettore X)
 - * ripetiamo per $i=1, \dots, r$:
 - * campioniamo un vettore $X^{(i)}=(X_1^{(i)}, \dots, X_k^{(i)})$ (da uno dei pt della griglia)
 - * ripetiamo per $j=1, \dots, k$ (componente per componente)
 - * calcoliamo $EE_j^{(i)}$ (elementary effect, incremental ratio), cioè quanto cambia la funzione f quando si sposta da $X^{(i)}$ nella direzione j di uno step nella griglia
 - * ripetiamo per $j=1, \dots, k$ (componente per componente)
 - * calcoliamo μ_j , ovvero la media di tutte le $|EE_j^{(i)}|$
 - * calcoliamo σ_j , ovvero la varianza delle $EE_j^{(i)}$

Per ogni direzione i (corrispondente alla comp. X_i) la media μ_i rappresenta the overall importance of X_i sull'output. (ogni componente X_i è tanto più importante quanto più μ_i è alto). La varianza σ_i è una misura della non-linearità / interaction effects. Se σ_i è alta allora X_i ha un'alta non-linearità o interagisce molto con le altre variabili.
- PROBLEMA: Curse of dimensionality. Selecting X, all the points that are related with the

evaluation of this OAT design will belong to the surface of the sphere centered in X of radius Δ (step of the grid). Il problema è che all'aumentare della dimensione in cui si trova X , il volume della sfera va rapidamente a zero (quindi, in a high dimensional space questo approccio è poco esplorativo e può far arrivare a risultati sbagliati)

GLOBAL (Variance Based SA)

- referred to as global SA because it considers tutto the variation range of the inputs
- Le derivate locali fanno un po' pena, bella la storia degli scatterplot MA:
Is it possible to introduce SA indicators that explain the variance of the model output rather than only the fraction of the variance associated with the linear surrogate model?
- L'idea è di introdurre SA indicators (SOBOL INDECES) that explain the variance of the output in terms of

* FIRST ORDER EFFECTS

- cosa succede alla varianza di Y se fissiamo il fattore X_i ?
Calcoliamo la varianza di Y condizionata a X_i fissata a un certo valore x_i^* :
 $\text{Var}(Y|X_i=x_i^*)$ è ovviamente minore di $\text{Var}(Y)$: è tanto più piccola quanto più X_i ha influenza sulla varianza totale di Y . Per far sì che questa misura non dipenda da x_i^* dobbiamo fare una media di $\text{Var}(Y|X_i=x_i^*)$, considerando tutti i x_i^* , cioè $E[.]$. According to the variance decomposition arriviamo a quello che definiamo first order effect di X_i su Y e FIRST ORDER SENSITIVITY INDEX S_i dell'input X_i su Y . Più S_i è alto e più X_i è importante (perché più S_i è alto e più è alta la % di varianza che possiamo spiegare con X_i).
- questo approccio è ottimo per FACTOR PRIORITIZATION (rank S_1, \dots, S_k in ordine decr.)
- buon approccio anche per factor fixing (fix the value of X_i for which $S_i < \epsilon$) dato un threshold ϵ
- nel caso di modelli lineari le S_i corrispondono con i SRC^2
- negli additive models (l'effetto totale degli input può essere espresso come la somma dei singoli effetti di ciascun input) la somma degli $S_i = 1$, se no < 1 (questo perché nei non additive model c'è della varianza che è spiegata dall'interazione dei termini, quindi mancano dei termini S_{ij}, S_{ijk}, \dots per poter arrivare a 1) \rightarrow higher order interactions aggiungendo i higher-order interactions we have found a way to recover (and so, understand) 100% of the variance of Y even for a nonadditive model.
Il problema è che se consideriamo tutte le possibili interazioni otteniamo $2^k - 1$ termini, too much.

* TOTAL EFFECTS

- per evitare di calcolare $2^k - 1$ termini introduciamo un altro indice: TOTAL EFFECT.
- Il total effect (o total sensitivity index) S_{T_i} di un input X_i su Y rappresenta una misura del contributo complessivo di X_i sull'output Y : questo indicatore tiene conto di tutte le interazioni possibili di X_i . (overall effect of a factor on the output)
- se $S_{T_i} = 0$ allora X_i è non influente: minore è S_{T_i} e meno X_i influenza il modello
- può essere usato nel FACTOR FIXING (in order to reduce the number of input parameters of a system, we can chop non-relevant parameters by fixing their values)
- una feature interessante è che questo metodo può calcolare il total effect di un input X_i oppure di un set di input X_i, \dots, X_j .
- Come mai questa decomposizione ha senso? Più che per una questione intuitivamente sensata? Supponiamo $Y=f(X)=f(X_1, \dots, X_k)$, dove le X_i sono indipendenti e uniformemente distribuite. La funzione $f(\cdot)$ può essere decomposta secondo la High-Dimensional Model Representation (HDMR), che è una functional analysis of variance (FANOVA) ed è una decomposizione finita e NON unica. Sotto ipotesi aggiuntive si arriva a una decomposizione unica di $f(\cdot)$, con tutti termini ortogonali tra di loro. Da qui si arriva alla decomposizione della $\text{Var}(Y)$.
- Tutto molto fico, ma per stimare questi indicatori abbiamo bisogno di molti run. Questo può essere un prob. In particolare: bisogna cover the input parameters space by samples and the perform a suitable MC approx of the expectations required to compute Sobol indeces (and total effect).
 \rightarrow Crude Monte Carlo è inefficiente, passiamo quindi a Quasi Monte Carlo.
- Esempio pratico e fico: Ishigami-Homma function
- E' possibile estendere the variance-based sensitivity analysis anche al caso dinamico (case of time dependent outputs)
 - definiamo una generalizzazione dei Sobol indeces (che tiene conto dell'intervallo di temp consid.)
 - i generalized indices misurano l'importanza dell'input tenendo conto della loro "storia" alla fine si ottiene quindi l'evoluzione dell'importanza degli input.
 - Esempio pratico: oscillatore

MOMENT-INDEPENDENT IMPORTANCE MEASURES

- tutto molto fico, ma vorremmo un metodo che non chiede di calcolare momenti (MC è pesante, anche se QMC)
- I moment independent importance measures sono dei sensitivity measures che take into account the change in the entire distribution of the model output (instead of the variation of one of its particular moments, e.g. variance)
- Ci sono due diversi moment-independent methods

* Δ SENSITIVITY MEASURE (basato sulla densità)

fissiamo la densità (incondizionata) dell'output Y , dopodiché riceviamo informazioni su una variabile X_i e decidiamo di fissarla a un valore x_i . Calcoliamo quanto è diversa la densità di $Y|X_i=x_i$ rispetto alla densità di Y per ogni possibile x_i e facciamo una media, ottenendo un valore Δ_i . Δ_i rappresenta the expected separation between the conditional and unconditional model output densities provoked by fixing X_i e viene chiamata Δ -sensitivity measure. Si ha che $\Delta_i=0$ se e solo se Y è indipendente da X_i . Più è alta Δ_i e più è alta la sensitivity di Y rispetto a X_i .
Notiamo però che questo metodo ha ancora come svantaggio che richiede l'approssimazione di diverse densità e valori medi.

* PAWN SENSITIVITY MEASURE (basato sulla funzione di ripartizione)

l'idea è di confrontare la funzione di ripartizione di Y (F_Y) e di $Y|X_i=x_i$ ($F_{Y|X_i=x_i}$). Calcoliamo una statistica T_i della distanza tra F_Y e $F_{Y|X_i=x_i}$ (ad esempio calcoliamo la distanza per ogni x_i e facciamo una media, oppure calcoliamo la distanza per ogni x_i e consideriamo distanza massima). Ovviamente possiamo variare sia la definizione di distanza tra le due funzioni di ripartizione (la più usata è la distanza di Kolmogorov-Smirnov) che la statistica T_i (massimo, media, ..). Comunque sia, X_i ha tanta più influenza su Y quanto più T_i è alta.

- Questo approccio può essere usato per factor prioritization and factor fixing.

- Nel caso di factor fixing si può fare ad esempio un two-sample KS test ($H_0: F_Y = F_Y|X_i$).
- Può essere applicato a qualsiasi input, anche time series (cosa che non possiamo fare con qualsiasi altro global SA)

SURROGATE MODELS or EMULATORS

- Fino ad ora abbiamo fatto:
Variance based SA -> Sobol indices (first order/total effects): there are moments to be computed (expectations, variances), it's very expensive -> Quasi Monte Carlo sampling on Sobol sequences
- usiamo Surrogate Models instead:
 - Polynomial chaos expansion
 - Gaussian Process regression
- Abbiamo visto che i global SA sono a powerful approach for determining the key random input parameters that drive the uncertainty of the model output predictions. Tuttavia sono troppo dispendiosi computazionali.
- E' un ottimo scenario per l'uso di modelli surrogati/emulatori
- Un modello surrogato è un'approssimazione del modello originale
- Alcuni metodi per creare surrogati sono polynomial chaos expansion (PCE) (da cui possiamo ricavare gli indici di Sobol analiticamente a partire dai coefficienti) e gaussian processes (GP) (da cui possiamo ricavare facilmente degli intervalli di confidenza per i sensitivity indices)

* POLYNOMIAL CHAOS EXPANSIONS

Consideriamo $Y=G(X)$, dove $X \sim f_X(\cdot)$. Assumendo che Y abbia varianza finita, è possibile usare la decomposizione spettrale $Y = \sum_j Z_j$ dove le Z_j sono un insieme numerabile di polinomi ortonormali (rispetto alla densità $f_X(\cdot)$), che creano una base, tale che $Z_j = \phi_j(\mathbf{X})$ e y_j sono i coefficienti (coordinate di Y nella nuova base). Truncating the expansion dopo N termini provides a finite-dimensional representation di qualsiasi funzione $G(X)$ of the random input X . Comunque si tratta di qualcosa di dispendioso siccome per calcolare le basi è necessario calcolare integrali. Inoltre è necessario conoscere la forma analitica di G .

- Computation of the coefficients:
Una volta che abbiamo fissato le basi (e deciso come troncare la serie) applichiamo un least-square minimization approach per trovare i coefficienti
- Sobol indices and PC expansion
La versione troncata della PC expansion contiene tutte le informazioni sulle proprietà statistiche della variabile aleatoria $Y=G(X)$. Both polynomial chaos expansions and Sobol decomposition (in the HDMR format) are sums of orthogonal functions. Taking advantage of this property, it is possible to derive analytic expressions for Sobol indices based on a PC expansion. For a given PC expansion, the Sobol indices at any order may be obtained by mere combination of the square of the coefficients.

* GAUSSIAN PROCESS REGRESSION

(Kriging). Si basa sulla conditional multivariate gaussian distribution. The principle of gaussian process regression is to consider that the prior knowledge about the model $G(X)$ can be modeled by a Gaussian process $Z(X)$ with a mean $\mu_Z(\cdot)$ and a covariance function (or kernel) C_Z . We consider that the true response is a realization of $Z(X)$. Starting from this prior distribution the goal is to determine the predictive distr. (for any possible x).

- Denotiamo \mathcal{X} a set of training samples e the corresponding model responses $\mathcal{Y}=G(\mathcal{X})$.
The goal is to perform out-of-sample prediction, that is to determine $Z(X_{\text{new}})$, updating the prior distribution $Z \sim GP(\cdot)$ incorporating the knowledge provided by the fact that $\mathcal{Y} = Z(\mathcal{X})$. The posterior distribution is then used to make predictions on unseen samples.

- Il modello è $Y=G(X)$
- Approssimiamo il modello come $Y=Z(X)$
- Partiamo da un set di dati $\{(x_1, y_1), \dots, (x_n, y_n)\}$, dove $y_i = G(x_i)$
- Costringiamo i dati a fittare l'approssimazione, cioè aggiorniamo media e varianza per far sì che $\{(x_1, y_1), \dots, (x_n, y_n)\} = \{(x_1, Z(x_1)), \dots, (x_n, Z(x_n))\}$
- Con la nuova media e varianza facciamo predizione

- Notiamo in particolare che abbiamo l'intera distribuzione degli output, non solo una predizione puntuale. Per cui possiamo ricavare intervalli di confidenza e stime dell'errore.
- The mean prediction is expressed as a linear combination of observations Y , plus the prior mean. The covariance prediction does not depend on the targets, only on the inputs
- Nel caso ci siano gaussian noise possiamo fare un modello un po' più loose (siccome prima avevamo la certezza di predizione nei punti già osservati, ora invece gli diamo un piccolo margine/intervallo)
- Media e covarianza della prior distribution sono iperparametri. Esistono diverse possibilità sia per la media (simple Kriging, ordinary kriging, universal kriging) che per la covarianza (Matern, lineare, esponenziale): diverse scelte portano a comportamenti diversi (in termini di smoothness o altro).
- Once the GP emulator is constructed, to perform a sensitivity analysis from such GP model two approaches are possible.
 - * possiamo sostituire il vero modello $G(X)$ con la media del conditional GP $\mu_Z|Y(\cdot)$ e fare SA sulla media. E' veloce ma produce un po' di bias, e in più non permette di quantificare l'errore
 - * possiamo sostituire il vero modello $G(X)$ con il GP $Z_N(X)$ avendo la distribuzione predittiva $Z(X)|Z(\mathcal{X})$. Con questo approccio possiamo quantificare l'errore.